AFRL-RW-EG-TP-2008-7403

DIMENSIONALITY REDUCTION AND INFORMATION-THEORETIC DIVERGENCE BETWEEN SETS OF LADAR IMAGES

David M. Gray
Air Force Research Laboratory
Munitions Directorate
AFRL/RWGI
Eglin AFB, FL 32542-6810

David M. Gray
José C. Principe
Computational NeuroEngineering Laboratory
University of Florida
P. O. Box 116130
Gainesville, FL 32611-6130

MARCH 2008

CONFERENCE PAPER

This paper was presented at the SPIE Defense and Security 2008 Conference, held in Orlando, Florida on 16-20 March 2008.

AIR FORCE RESEARCH LABORATORY, MUNITIONS DIRECTORATE

■ **Air Force Materiel Command** ■ **United States Air Force** ■ **Eglin Air Force Base**

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX 03 2008	CONFERENCE PAPER	

4. TITLE AND SUBTITLE

DIMENSIONALITY REDUCTION AND INFORMATION-THEORETIC DIVERGENCE BETWEEN SETS OF LADAR IMAGES

5a. CONTRACT NUMBER

N/A

5b. GRANT NUMBER

N/A

5c. PROGRAM ELEMENT NUMBER

61102F

6. AUTHOR(S)

1 & 2. David M. Gray
2. José C. Principe

5d. PROJECT NUMBER

2311

5e. TASK NUMBER

EM

5f. WORK UNIT NUMBER

04

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

1. Air Force Research Laboratory
Munitions Directorate
AFRL/RWGI
Eglin AFB, FL 32542-6810

2. Computational NeuroEngineering Laboratory
University of Florida
P. O. Box 116130
Gainesville, FL 32611-6130

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-RW-EG-TP-2008-7403

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Munitions Directorate
AFRL/RWGI
Eglin AFB, FL 32542-6810

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL-RW-EG

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

SAME AS BLOCK 8

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION A: Approved for public release; distribution unlimited. 96ABW/PA Public Release Approval Confirmation #02-28-08-107; dated 28 February 2008.

13. SUPPLEMENTARY NOTES

See 'cover page' for pertinent metadata information.

14. ABSTRACT

This paper presents a preliminary study of information-theoretic divergence between sets of LADAR image data. This study has been motivated by the hypothesis that despite the huge dimensionality of raw image space, related images actually lie on embedded manifolds within this set of all possible images and can be represented in much lower-dimensional sub-spaces. If these low-dimensional representations can be found, information theoretic properties of the images can be exploited while circumventing many of the problems associated with the so-called “curse of dimensionality.” In this study, PCA techniques are used to find a low-dimensional sub-space representation of LADAR image sets. A real LADAR image data set was collected using the AFSTAR sensor and a synthetic image data set was created using the Irma LADAR image modeling program. One unique aspect of this study is the use of an entirely synthetic data set to find a sub-space representation that is reasonably valid for both the synthetic data set and the real data set. After the sub-space representation is found, an information-theoretic density divergence measure (Cauchy-Schwarz divergence) is computed using Parzen window estimation methods to find the divergence between and among the sets of synthetic and real target classes. These divergence measures can then be used to make target classification decisions for sets of images. In practice, this technique could be used to make classification decisions on multiple images collected from a moving sensor platform or from a geographically distributed set of cooperating sensor platforms operating in a target region.

15. SUBJECT TERMS

Keywords: information-theoretic learning; image dimensionality reduction; PCA; ATR; information theory; LADAR; cooperative sensors; Cauchy-Schwarz divergence; Parzen window

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			David Gray
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	13	19b. TELEPHONE NUMBER *(Include area code)*

Dimensionality reduction and information-theoretic divergence between sets of LADAR images

David M. Gray[*ab] and José C. Príncipe[a]
[a]Computational NeuroEngineering Laboratory, University of Florida, P.O. Box 116130, Gainesville, FL USA 32611-6130;
[b]Air Force Research Laboratory, 101 W. Eglin Blvd., Suite 210, Eglin AFB, FL USA 32542-6810

ABSTRACT

This paper presents a preliminary study of information-theoretic divergence between sets of LADAR image data. This study has been motivated by the hypothesis that despite the huge dimensionality of raw image space, related images actually lie on embedded manifolds within this set of all possible images and can be represented in much lower-dimensional sub-spaces. If these low-dimensional representations can be found, information theoretic properties of the images can be exploited while circumventing many of the problems associated with the so-called "curse of dimensionality." In this study, PCA techniques are used to find a low-dimensional sub-space representation of LADAR image sets. A real LADAR image data set was collected using the AFSTAR sensor and a synthetic image data set was created using the Irma LADAR image modeling program. One unique aspect of this study is the use of an entirely synthetic data set to find a sub-space representation that is reasonably valid for both the synthetic data set and the real data set. After the sub-space representation is found, an information-theoretic density divergence measure (Cauchy-Schwarz divergence) is computed using Parzen window estimation methods to find the divergence between and among the sets of synthetic and real target classes. These divergence measures can then be used to make target classification decisions for sets of images. In practice, this technique could be used to make classification decisions on multiple images collected from a moving sensor platform or from a geographically distributed set of cooperating sensor platforms operating in a target region.

Keywords: information-theoretic learning; image dimensionality reduction; PCA; ATR; information theory; LADAR; cooperative sensors; Cauchy-Schwarz divergence; Parzen window

1. INTRODUCTION

In recent years there has been considerable development and progressive maturation of a new statistical machine-learning paradigm which has been coined "information theoretic learning" (ITL) by Principe, et al [1]. ITL incorporates information theoretic cost functions and is therefore able to utilize statistical relationships in the data beyond the common second-order correlation. During roughly the same time-frame, there has been significant interest in the issue of dimensionality reduction, particularly in how it relates to dealing with the high dimensionality of image data. The prospect for effective image dimensionality reduction techniques is motivated by the belief that despite the high dimensionality of raw image data (number of dimensions $= ncols \bullet nrows$), related images can be considered to lie on low-dimensional manifolds embedded in the high-dimensional image space. There has been considerable development of linear dimensionality reduction techniques, e.g. principle component analysis, independent component analysis (ICA), and multidimensional scaling (MDS), etc. and, more recently, non-linear dimensionality reduction techniques, e.g. kernel PCA, Isomap, locally linear embedding (LLE), and local tangent space alignment (LTSA), etc [2].

This paper outlines the preliminary results of research investigating information-theoretic divergence measures applied to laser radar (LADAR) images of five different objects (see Fig. 1) viewed at a fixed depression angle from numerous aspect angles around the object. Due to the relatively high range and angular resolution of LADAR sensors, this data

[*]david.gray@eglin.af.mil; phone 1 850 883-0849; fax 1 850 882-3344

provides a representation of the object as a function of viewing geometry and the object's inherent shape. In this paper, we treat the collection of all aspect views of one object as a single "class" and investigate the divergence between the five classes of targets resulting from the collection of views of each of the five objects. Estimation of the divergence measure requires estimation of the class probability density functions (pdfs) which is normally difficult due to the high dimensionality of the image data and the inter-related issue of insufficient number of data points to adequately sample the space. We address these issues by 1) supplementing real data with synthetic data, and 2) performing dimensionality reduction on the data to reduce the impact of the "curse of dimensionality." One unique aspect of this research is the use of exclusively synthetic imagery to find a subspace representation (dimensionality reduction) that is reasonably valid for the real data. The rest of this paper is organized as follows. Section 2 describes the data used in the research. Section 3 briefly describes the PCA-based dimensionality reduction technique and addresses the issue of how many dimensions to use in the low-dimensional representation. Section 4 provides some visualizations of the low-dimensional data representations. Section 5 discusses the Cauchy-Schwarz information theoretic divergence measure, its estimation using Parzen window techniques, and briefly discusses the issue of Parzen window size. Section 6 presents the results of the divergence estimation between classes using both the synthetic and real data. Section 7 presents a summary of the paper and brief conclusions from the results. Finally, Section 8 outlines future research directions that have been motivated by the current work.

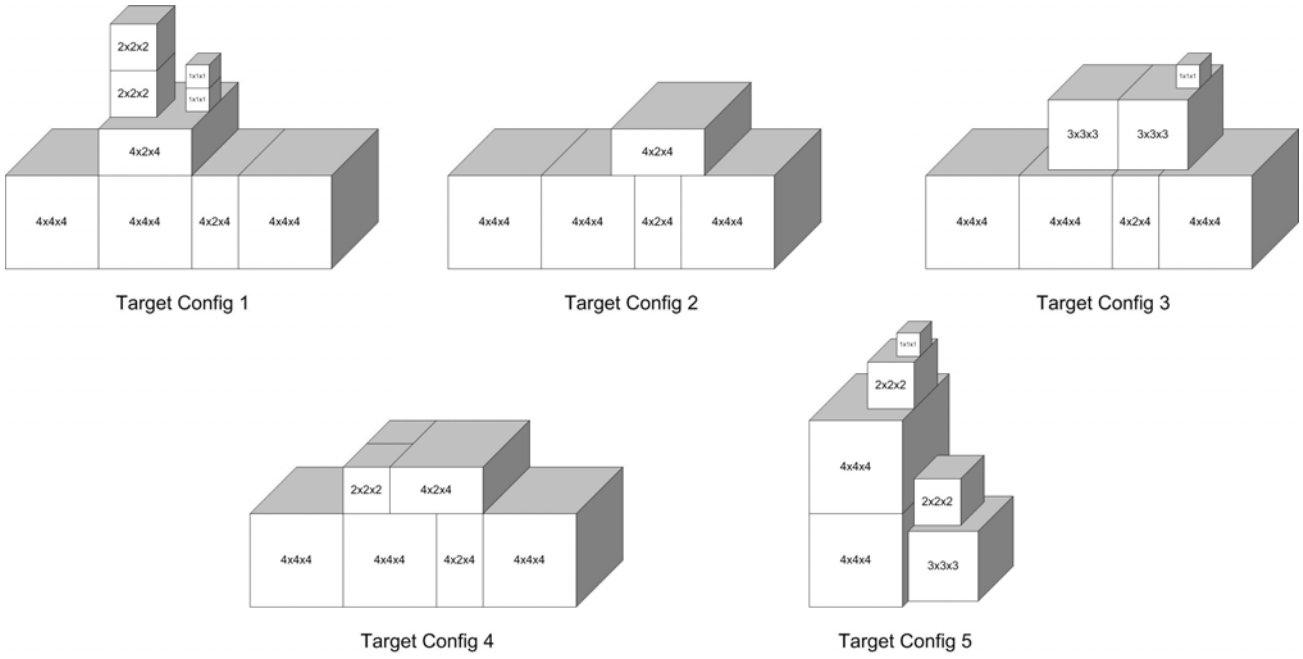


Fig. 1. Diagrams of the different target configurations used in the data collection.

2. DATA

For this research, both real and synthetic data were utilized. Real LADAR images of the target objects were collected using the AFSTAR sensor while synthetic LADAR images of the target objects were generated using the Irma modeling software.

2.1 Real Data Collection

The intent of this collection effort was to create a database of LADAR images that have been collected under relatively well-controlled conditions and with accurately known physical objects matching the desired five target configurations. Toward this end, a number of target "boxes" were specifically built for deployment during the data collection efforts. These "boxes" are fabricated from aluminum sheets and have been constructed in various sizes (1x1x1, 2x2x2, 3x3x3,

4x4x4, and 4x4x2, all dimensions given in feet.) For purposes of the test, the various target configurations were created by stacking the individual boxes to form the composite targets.

The data collection was conducted at the Russell Measurement Facility at Redstone Arsenal in Huntsville, Alabama. This facility consists of a 300 foot high tower, in which the sensor was located, and surrounding test fields where the desired targets can be positioned at various distances from the sensor and in various background conditions. Additionally, a target "turntable" is available for use during the collection efforts. The turntable consists of a mobile platform upon which desired targets can be placed and then rotated through various target aspect angles via remote computer control.

The AFSTAR LADAR sensor was used to collect the real images used in this research effort. The AFSTAR sensor was developed as a breadboard sensor to demonstrate the feasibility of the laser radar imaging modality for munitions-based ATR applications. A significant aspect of this demonstration has been the collection of a large database of images for the development of ATR algorithms. Significant design and operating parameters of the AFSTAR sensor are summarized in Table 1. A photograph of the AFSTAR sensor is shown in Fig. 2.

Table 1. AFSTAR Design and Operating Parameters.

Parameter	**Value**
Detection Mode	Direct Detection
Wavelength	$1.06 \mu m$
Image Size	148 rows by 301 columns

In preparation for this collection effort, the target boxes were painted with white paint and "dusted" with reflective glass beads. These beads are the type used in traffic marking paint to provide high visibility during night and adverse weather conditions since the glass beads act as tiny retro-reflectors of automobile headlights. These beads also proved to have superior reflective performance at the laser operating wavelength of 1.06 microns and were used to "dust" the surface of the target boxes to improve the retro-reflective return of the laser radar, thereby reducing pixel dropouts.

During this collection effort, LADAR images of target configurations 1,2,3,4, and 5 were collected at a slant range distance of approximately 250 meters. Again, data was collected at a fixed depression angle corresponding to that presented to a sensor located at an altitude of 300 feet, looking toward a target located at a horizontal distance of approximately 233 meters. Data was collected at azimuth angles from 0 to 180 degrees at 5 degree increments. Table 2 compiles the pertinent collection parameters and test points that comprise this particular data set. A representative photograph of the data collection setup is shown in Fig 3.

Table 2. Data Collection Specifications.

Fixed Parameters	**Value**
Nominal Sensor Tower Position	300 feet
Nominal Slant Range to Target	250 meters
Nominal Horizontal Range to Target	233 meters
Nominal Sensor Depression Angle	21.5°
Target Paint	White w/ reflective glass beads
Variable Parameters	**Values**
Target Configuration	1,2,3,4, and 5
Target Aspect	0° - 180°, 5° spacing

2.2 Irma Modeling Software

The Irma model is a computer software code for synthetic scene generation developed by the Air Force Research Laboratory (AFRL). Irma is capable of generating co-registered synthetic scenes in 1) passive infrared, 2) passive millimeter wave, 3) active infrared (laser radar), and 4) active millimeter wave. For the purposes of this study, only the laser radar channel was utilized. The Irma model uses geometric descriptions, or CAD models, of the targets and backgrounds from which to render the synthetic scenes. The geometric descriptions consist of triangular facets and ellipsoids or quadrics. Flat target regions are represented by a smaller number of relatively large triangular facets or ellipsoids while curved regions are represented by a larger number of relatively small triangular facets or ellipsoids. Irma is capable of rendering synthetic images as would be collected using specific sensor hardware. Sensor hardware

parameters are entered in the model to describe the sensor. Scanning and staring sensor systems can be modeled in significant detail. The Irma output consists of a rectangular grid of pixels representing the image of the scene that would be collected by the modeled sensor. The LADAR channel of Irma allows for the generation of both high-fidelity and medium-fidelity synthetic imagery of monostatic ranging (time-of-flight) LADAR systems. The high-fidelity mode creates scenes that are useful for evaluating sensor and scene specific parameters necessary for engineering trade studies. The medium-resolution mode is used when generating large quantities of synthetic LADAR imagery that is required for algorithm development and assessment.

2.3 Representative Data

Representative synthetic LADAR images generated by Irma are shown in Fig. 4. Each target configuration is shown at 90° aspect (broadside). Representative real LADAR images collected by AFSTAR are shown in Fig. 5. Again, each target configuration is shown at 90° aspect (broadside).

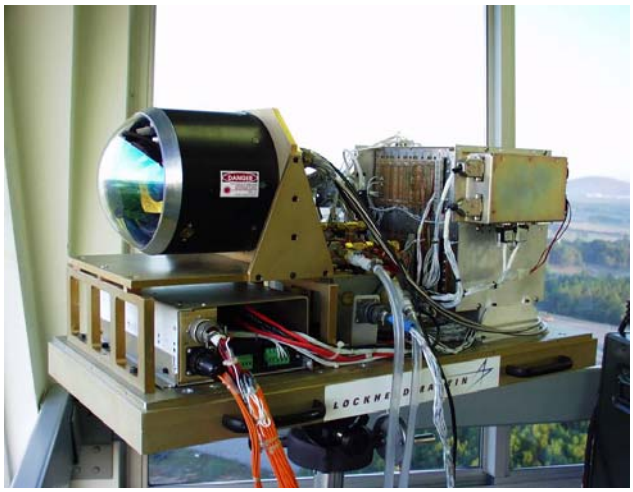


Fig. 2. The AFSTAR LADAR sensor.

Fig. 3. The data collection setup showing target configuration 1 on turntable at 90° aspect (broadside).

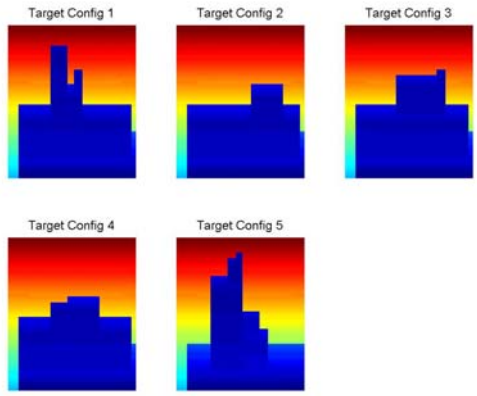


Fig. 4. Synthetic (Irma) LADAR images of each target configuration at 90° aspect (broadside).

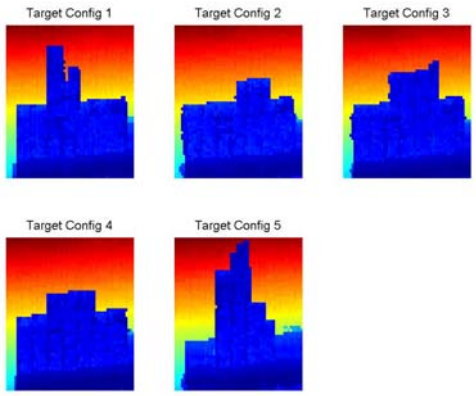


Fig. 5. Real (AFSTAR) LADAR images of each target configuration at 90° aspect (broadside).

3. DIMENSIONALITY REDUCTION

3.1 PCA-based Dimensionality Reduction

For the results reported herein, the well-known Principle Component Analysis (PCA) was used to perform the dimensionality reduction. In their paper on face recognition [3], Turk and Pentland showed that PCA could be efficiently computed for a set of M images of size $N = nrows \cdot ncols$ since the rank of the correlation matrix is M rather than N. The derivation will not be repeated here for space reasons, but can be found in the reference. If the eigenvectors obtained by PCA are arranged in descending order according to the magnitude of the corresponding eigenvalue, the relative contribution of each eigenvector toward the variance of the set of images is obtained. This ordered representation of the eigenvalues is referred to as the eigenspectrum and is shown for the synthetic data set in Fig 6. This data has the typical eigenspectrum plot with a small number of components accounting for the majority of the variance. The dimensionality reduction is achieved by retaining only k of the M (where, typically, $k << M$) eigenvector components in the representation.

3.2 Choosing the number of reduced dimensions – intrinsic dimensionality

One issue in dimensionality reduction is choosing the appropriate number of dimensions (the k) to include in the reduced dimensionality data representation. Ideally, this dimension would correspond to the so-called intrinsic dimensionality of the data [4]. A number of somewhat principled methods have been proposed for finding the intrinsic dimensionality [5] [6], but have not been investigated here. For this paper, $k = 3$ was chosen somewhat arbitrarily based on the classification results obtained from a nearest neighbor classifier. Table 3 presents the confusion matrix for the synthetic images when represented with the first 3 PCA dimensions. In this classifier, the declared class of the even degree {2°, 4°, 6°, . . ., 360°} aspect images is the same as the actual class of the nearest (Euclidean distance) odd degree {1°, 3°, 5°, . . ., 359°} aspect image. Table 4 presents similar results for the real images when projected into the subspace found using PCA on the synthetic images. In this case, the declared class of image aspects evenly divisible by 10°, i.e. {0°, 10°, 20°, . . ., 180°} is the same as the actual class of the nearest image from the set {5°, 15°, 25°, . . ., 175°}. For comparison, the same nearest neighbor classification of the real images yields only 97.9% overall accuracy (an improvement of only 5.3%) when the dimensionality is not reduced at all, i.e. data is represented with all N dimensions.

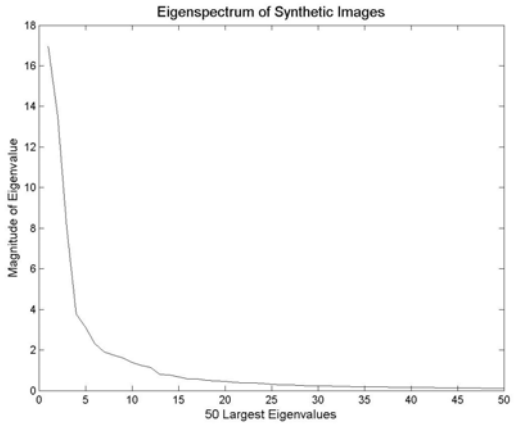


Fig. 6. Eigenspectrum of the synthetic images showing the 50 largest eigenvalues.

Table 3. Confusion Matrix when images are represented with 3 PCA dimensions. Nearest neighbor classifier – even synthetic to odd synthetic.

	1	2	3	4	5	%
1	180	0	0	0	0	100
2	0	180	0	0	0	100
3	0	0	180	0	0	100
4	0	0	0	180	0	100
5	0	0	0	0	180	100
Overall						100

Table 4. Confusion Matrix when images are represented with 3 PCA dimensions. Nearest neighbor classifier – Even real to odd real.

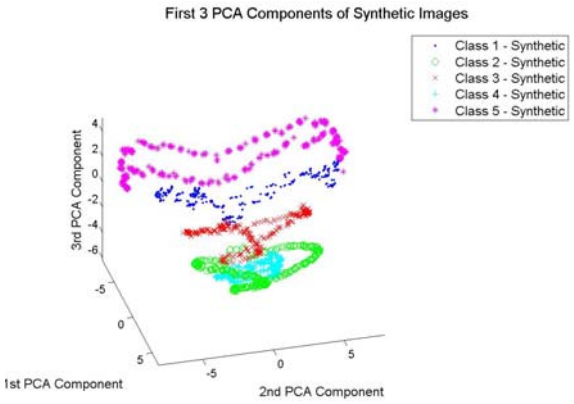
	1	2	3	4	5	%
1	18	0	0	0	1	94.7
2	0	17	0	2	0	89.5
3	1	0	17	1	0	89.5
4	0	2	0	17	0	89.5
5	0	0	0	0	19	100.0
Overall						92.6

3.3 Use of Synthetic Data to Find Subspace Projection for Dimensionality of Real Data

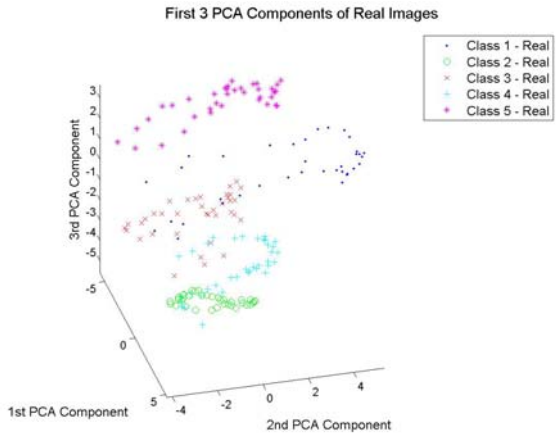
We point out again that one unique aspect of the research presented here is the use of synthetic LADAR data to compute the PCA dimensionality reduction subspace projection, and subsequently applying this subspace projection to reduce the dimensionality of the real data. As discussed above, applying a nearest neighbor classifier on the real data when projected into the subspace found using the synthetic data resulted in an overall classification accuracy of 92.6 %. Applying the same nearest neighbor classifier on the real data when projected into the subspace found using the real data resulted in an overall classification accuracy of only 89.5%. Although probably not a significant statistical difference, it appears that the subspace projection found using the synthetic data is reasonably valid for the real data.

4. DIMENSIONALITY REDUCTION RESULTS AND VISUALIZATION

After the dimensionality reduction has been applied, it is interesting to visualize the results. Since we have used $k = 3$ it is easy to visualize the reduced data sets using 3-D plots. These visualizations are shown in different ways in videos 1 – 6. Video 1 shows the first 3 principle components for each aspect of each target (Class) of the synthetic images. Video 2 shows the first 3 principle components for each aspect angle of each target for the real images. Separation is not perfect but the general clustering trend is obvious. Note that Class 2 and Class 4 (very similar physically) lie relatively close together in the subspace projection – an intuitively pleasing result.

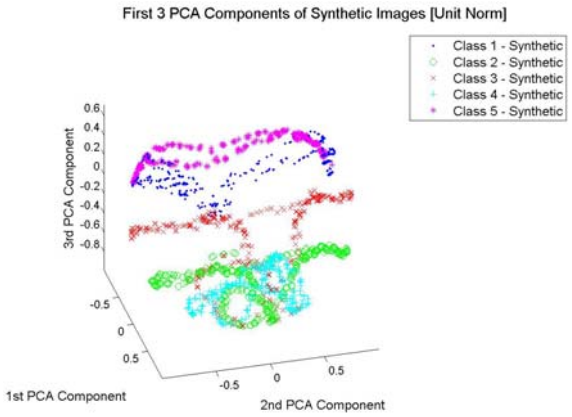


Video 1. Visualization of the first 3 PCA components for the synthetic images.
http://dx.doi.org/doi.number.goes.here

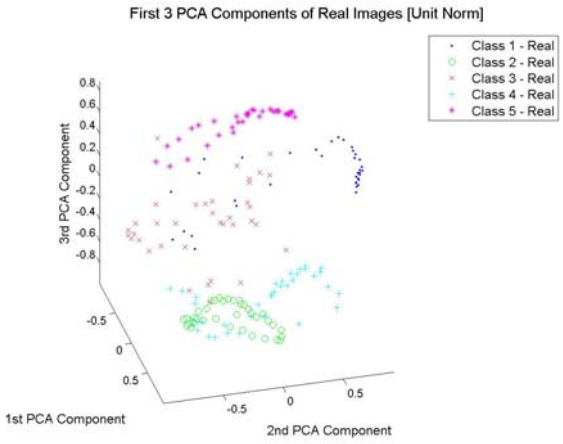


Video 2. Visualization of the first 3 PCA components for the real images.
http://dx.doi.org/doi.number.goes.here

Videos 3 and 4 show the corresponding data after each point has been normalized to lie on the unit circle. The intent here is to observe effects after removal of the scale variation between the synthetic data and the real data.

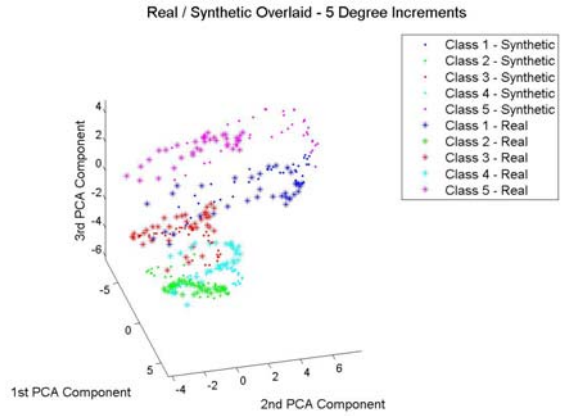


Video 3. Visualization of the first 3 PCA components [Unit Norm] for the synthetic images. http://dx.doi.org/doi.number.goes.here

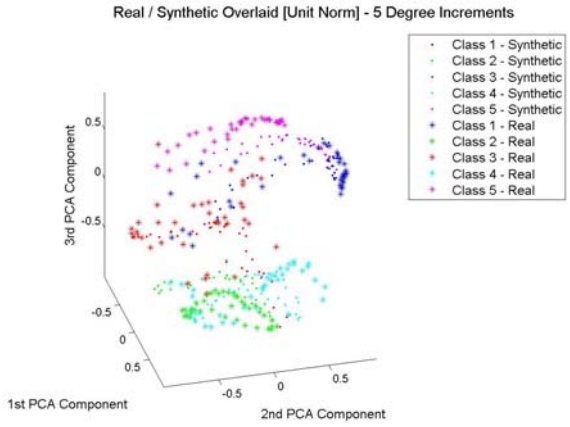


Video 4. Visualization of the first 3 PCA components [Unit Norm] for the real images. http://dx.doi.org/doi.number.goes.here

Finally, videos 5 and 6 show the real data and synthetic data overlaid on the same plot. Here the synthetic data is plotted only at aspect angles at which the real data was collected (i.e. 0°, 5°, 10°, . . ., 180°). Video 5 shows the original data and video 6 shows the data normalized to the unit circle.



Video 5. Visualization of the first 3 PCA components – Synthetic and real images overlaid – Multiples of 5° in aspect. http://dx.doi.org/doi.number.goes.here



Video 6. Visualization of the first 3 PCA components [Unit Norm] – Synthetic and real images overlaid – Multiples of 5° in aspect. http://dx.doi.org/doi.number.goes.here

5. INFORMATION-THEORETIC DIVERGENCE

5.1 Cauchy-Schwarz Divergence

In our research we are motivated to apply information-theoretic divergence measures to our data since the measures provide an indication of how "close" one pdf lies to another. The foundation for these divergence measures is information entropy and a number of such entropy measures exist (e.g. Shannon's entropy [7] and Renyi's entropy [8]). There are a number of divergence measures based on the various entropy definitions. For our purposes, we choose the Cauchy-Schwarz divergence measure because of some attractive properties that can be applied in conjunction with

Parzen window approximation techniques. Following the derivation in [9], the Cauchy-Schwarz (CS) divergence is given by

$$D_{CS} = -\log\left[\frac{\int p(\mathbf{x})q(\mathbf{x})d\mathbf{x}}{\sqrt{\int p^2(\mathbf{x})d\mathbf{x}\int q^2(\mathbf{x})d\mathbf{x}}}\right].$$

They estimate D_{CS} by using Parzen windowing techniques to estimate the pdf's, $p(\mathbf{x})$ and $q(\mathbf{x})$. Now the estimates for $p(\mathbf{x})$ and $q(\mathbf{x})$ be given by $\hat{p}(\mathbf{x})$ and $\hat{q}(\mathbf{x})$ respectively. Thus, the Parzen window estimates of the pdfs are

$$\hat{p}(\mathbf{x}) = \frac{1}{N_p}\sum_{i=1}^{N_p} W_h(\mathbf{x},\mathbf{x}_i), \quad \hat{q}(\mathbf{x}) = \frac{1}{N_q}\sum_{j=1}^{N_q} W_h(\mathbf{x},\mathbf{x}_j)$$

where $W(\cdot)$ is the window function and h is the window size or "bandwidth" parameter.

The window function must integrate to unity so it is typically chosen to be a pdf functional form such as the Gaussian kernel, the Epanechinikov, triangle, uniform, bi-weight, or tri-weight, etc. Continuing to follow their development, we have chosen to use the d-dimensional Gaussian kernel given by

$$G_{\sigma^2}(\mathbf{x},\mathbf{x}_i) = \frac{1}{(2\pi\sigma^2)^{d/2}}\exp\left\{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma^2}\right\}$$

because of the convenient convolution theorem for Gaussian functions which states that

$$\int G_{\sigma^2}(\mathbf{x},\mathbf{x}_i)G_{\sigma^2}(\mathbf{x},\mathbf{x}_j)d\mathbf{x} = G_{(\sqrt{2}\sigma)^2}(\mathbf{x}_i,\mathbf{x}_j).$$

Using this property, we then obtain our estimate for the Cauchy-Schwarz divergence,

$$\hat{D}_{CS}(p,q) = -\log\frac{\frac{1}{N_pN_q}\sum_{i=1}^{N_p}\sum_{j=1}^{N_q} G_{2\sigma^2}(\mathbf{x}_i,\mathbf{x}_j)}{\sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}\sum_{i'=1}^{N_p} G_{2\sigma^2}(\mathbf{x}_i,\mathbf{x}_{i'})\frac{1}{N_q}\sum_{j=1}^{N_q}\sum_{j'=1}^{N_q} G_{2\sigma^2}(\mathbf{x}_j,\mathbf{x}_{j'})}}$$

5.2 Window Size Parameter

The choice of window size, rather than window type, has generally been shown to be the most influential in the success of the Parzen window pdf estimation technique. In [10], Silverman showed that the optimum window size depends on the data itself, but some general guidelines for the window size have been developed. For data of dimension d, Silverman introduced a widely-adopted guideline "Silverman's rule-of-thumb - SROT" for the window size selection, given by

$$h_{\text{SROT}} = 0.9Ad^{-1/5}.$$

Various slight changes in the value of the constant and for the parameter A have been proposed. For example, in [11], DiNarclo and Tobias use

$$A = \min(\text{sample standard deviation, (sample interquartile range/1.34)}).$$

Silverman's rule-of-thumb for the Parzen window size was used for the results reported in this paper.

6. INFORMATION-THEORETIC DIVERGENCE RESULTS

Results of the divergence calculations using the estimation techniques of Section 5 are shown in Tables 5 – 9. Table 5 shows the divergence estimate between the 5 classes using the synthetic data for both p and q. Note, as expected, that the estimates are symmetric and equal to zero when p = q. Table 6 shows the divergence estimate between the 5 classes when the even degree aspect data is used for p and the odd degree aspect data is used for q. Again, note that the divergence measure is approximately symmetric and near zero when the class of p and q are both the same. The same trends can be seen in Table 7 which shows the divergence estimate between the 5 classes when the real data is used for both p and q. The same general trends are also observed in Table 8 which shows the divergence estimates when the data from alternating aspect angles are used for p and q. Finally, Table 9 shows the divergence estimates when the synthetic data is used for p and the real data is used for q. Although some significant variations exist, especially with regard to symmetry, generally similar trends for the divergence between classes can be observed. Notice throughout these calculations that the divergence estimate between classes 2 and 4 generally indicates that these classes lie relatively close to one another, i.e. the divergence measure is small. This is intuitively pleasing since the two objects are physically very similar. Note, in general, that it would be possible to make classification decisions between the various sets of images based on the estimated divergence measures.

Table 5. Cauchy-Schwarz Divergence. Synthetic to synthetic.

	1	**2**	**3**	**4**	**5**
1	0	3.6809	1.5269	4.9862	2.0431
2	3.6809	0	1.1362	0.3284	8.6701
3	1.5269	1.1362	0	1.0699	5.3037
4	4.9862	0.3284	1.0699	0	11.7697
5	2.0431	8.6701	5.3037	11.7697	0

Table 6. Cauchy-Schwarz Divergence. Even synthetics to odd synthetics.

	1	**2**	**3**	**4**	**5**
1	0.0004	3.6842	1.5316	4.9482	2.0525
2	3.6796	0.0002	1.1349	0.3301	8.6305
3	1.5231	1.1380	0.0003	1.0689	5.3000
4	5.0268	0.3269	1.0712	0.0001	11.7764
5	2.0344	8.7138	5.3095	11.7645	0.0004

Table 7. Cauchy-Schwarz Divergence. Real to real.

	1	**2**	**3**	**4**	**5**
1	0	4.4974	2.1571	2.8381	2.0241
2	4.4974	0	1.9718	0.5232	11.2804
3	2.1571	1.9718	0	1.1798	5.1102
4	2.8381	0.5232	1.1798	0	7.3966
5	2.0241	11.2804	5.1102	7.3966	0

Table 8. Cauchy-Schwarz Divergence. Even real to odd real.

	1	2	3	4	5
1	0.0074	4.4928	2.1269	2.8175	2.0146
2	4.5274	0.0051	1.9529	0.5259	11.0962
3	2.2062	2.0016	0.0096	1.1985	5.0484
4	2.8673	0.5293	1.1705	0.0056	7.2742
5	2.0525	11.5580	5.2233	7.5556	0.0072

Table 9. Cauchy-Schwarz Divergence. Synthetic (multiples of 5°) to real (multiples of 5°).

	1	2	3	4	5
1	0.1829	9.5980	4.5099	7.0863	2.5522
2	4.2893	0.3463	1.9084	0.7077	10.8308
3	2.9048	2.2187	0.2898	1.8164	6.0469
4	5.5519	0.4845	2.1805	0.1753	14.1852
5	1.2137	16.3793	8.5352	13.7688	0.8491

7. SUMMARY AND CONCLUSIONS

In this paper we have presented the results of an information-theoretic divergence measure between sets of LADAR images of target-like objects where the members of each class are different views of the target object. Initially, a PCA-based dimensionality reduction method was used to find a low-dimensional manifold representation of the high-dimensional LADAR image data and thereby overcome the problems associated with high-dimensional pdf estimation caused by the "curse of dimensionality." Importantly, the subspace projection calculated from synthetic data was found to be somewhat reasonable to use for the subspace projection of the real data. 3-D visualizations of the low-dimensional data were presented to provide insight into the clustering of the data. After finding the low-dimensional data representation, an information-theoretic divergence measure was estimated using non-parametric Parzen windowing methods. Although further work needs to be performed to assess the quality of the divergence estimates, there are a number of intuitively pleasing aspects of the calculated estimates that subjectively support the belief that they are reasonable, i.e.

a) The estimates generally tend to follow the symmetry property of the Cauchy-Schwarz divergence, i.e.

$$D_{CS}(p, q) = D_{CS}(q, p)$$

b) The estimates generally tend to obey another property of the Cauchy-Schwarz divergence, i.e.

$$D_{CS}(p, q) = 0 \; iff \; p = q$$

c) Images that are similar in "shape" seem to have a small Cauchy-Schwarz divergence.

8. FUTURE WORK

This preliminary research has identified a number of fertile areas in which to concentrate our future research. Two immediately identifiable areas are the investigation of more complex dimensionality reduction techniques (both linear and non-linear) and a more sophisticated exploration of the Parzen window size in the divergence calculation. As previously mentioned, for the currently reported results the dimensionality reduction was performed using PCA. In the future, results will be reported (if the technique is applicable) for kernel PCA, local linear embedding, multidimensional scaling, etc. In the current research, the divergence calculations were performed using Silverman's rule-of-thumb to set a single window width parameter for all dimensions. Future work will focus on a more thorough exploration of the effect of the window width parameter, cross-validation methods, other adaptive techniques, and non-symmetric Gaussian kernels. Another area for future research is to investigate ways in which the synthetic data might be transformed to better match the real data in the subspace projection. Successful implementation of this concept could make synthetic data more useful for augmenting ATR training and testing databases that always seem to have limited (or non-existent) real data.

REFERENCES

[1] Principe, J., Xu, D., and Fisher, J., "Information theoretic learning," in: S. Haykin (Ed.), Unsupervised Adaptive Filtering, vol. I, Wiley, New York, 2000.

[2] Zhang, T., Yang, J., Zhao, D., and Ge, X., "Linear local tangent space alignment and application to face recognition," Neurocomputing 70 (2007) 1547-1553.

[3] Turk, M., and Pentland, A., “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, March, 1991.
[4] van der Maaten, L.J.P., Report MICC 07-07, “An Introduction to Dimensionality Reduction Using Matlab,” Universiteit Maastricht, July 2007.
[5] Erdogmus, D., Hild, K., and Principe, J., “Kernel Size Selection in Parzen Density Estimation,” submitted to Journal of VLSI Signal Processing-Systems.
[6] Comaniciu, D., “An Algorithm for Data-Driven Bandwidth Selection,” IEEE PAMI, vol. 25, no.2, Feb. 2003,pp. 281-288.
[7] Shannon, C., “A mathematical theory of communication,” Bell Syst. Tech. J. 27 (1948) 379-423 623-653.
[8] Renyi, A., On measures of entropy and information, Selected Papers of Alfred Renyi, vol. 2, Akademiai Kiado, Budapest, 1976, pp. 565-580.
[9] Jenssen, R., Principe, J., Erdogmus, D., and Eltoft, T., “The Cauchy-Schwarz divergence and Parzen windowing: Connections to graph theory and Mercer kernels,” Journal of the Franklin Institute, 343 (2006) 614-629.
[10] Silverman, B., Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
[11] DiNarclo, J., and Tobias, J., “Nonparametric Density and Regression Estimation,” The Journal of Economic Perspectives, vol, 15, no. 4, 2001, pp. 11-28.